

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: LaVALLIE, EDWARD  
RACIE, LISA
- (ii) TITLE OF INVENTION: HUMAN SDF-5 PROTEIN AND COMPOSITIONS
- (iii) NUMBER OF SEQUENCES: 3
- (iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: GENETICS INSTITUTE, INC.  
(B) STREET: 87 CAMBRIDGE PARK DRIVE  
(C) CITY: CAMBRIDGE  
(D) STATE: MA  
(E) COUNTRY: USA  
(F) ZIP: 02140
- (v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:  
(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: LAZAR, STEVEN R.  
(B) REGISTRATION NUMBER: 32,618
- (ix) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: (617) 498-8260  
(B) TELEFAX: (617) 876-5851

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2027 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCCGCC TTCATGGCCT AGCTCATTCT GCTCCCCCGG GTGGAGCCC CCCGGAGCTG 60  
CGCGCGGGCT TGCAGCGCCT CGCCCGCGCT CCTCCCGGTG TCCCGCTTCT CCGCGCCCCA 120  
GCCGCCGGCT GCCAGCTTTT CGGGGCCCCG AGTCGCACCC AGCGAAGAGA GCAGGGCCCGG 180  
GACAAGCTCG AACTCCGGCC GCCTCGCCCT TCCCCGGCTC CGCTCCCTCT GCCCCCTCGG 240  
GGTCGCGCGC CCACGATGCT GCAGGGCCCT GGCTCGCTGC TGCTGCTCTT CCTCGCCCTCG 300  
CACTGCTGCC TGGGCTCGGC GCGCGGGCTC TTCCCTCTTG GCCAGCCCGA CTTGTCCTAC 360  
AAGCGCAGCA ATTGCAAGCC CATCCCGGCC AACCTGCAGC TGTGCCACGG CATCGAAATAC 420

2027 1980 1950 1920 1890 1860 1830 1800 1770 1740 1710 1680 1650 1620 1590 1560 1530 1500 1470 1440 1410 1380 1350 1320 1290 1260 1230 1200 1170 1140 1110 1080 1050 1020 990 960 930 900 870 840 810 780 750 720 690 660 630 600 570 540 510 480

CAGAACATGC GGCTGCCAA CCTGCTGGC CACGAGACCA TGAAGGAGGT GCTGGAGCAG 480  
 GCCGGCGCTT GGATCCCGCT GGTCACTGAAG CAGTGCCACC CGGACACCAA GAAGTTCCCTG 540  
 TGCTCGCTCT TCGCCCCCGT CTGCCTCGAT GACCTAGACG AGACCATCCA GCCATGCCAC 600  
 TCGCTCTGCG TGGAGGTGAA GGACCGCTGC GCCCCGGTCA TGTCCGCCTT CGGCTTCCCC 660  
 TGGCCCGACA TGCTTGAGTG CGACCGTTTC CCCCAGGACA ACGACCTTG CATCCCCCTC 720  
 GCTAGCAGCG ACCACCGTCCCT GCCAGCCACC GAGGAAGCTC CAAAGGTATG TGAAGCCTGC 780  
 AAAAATAAAA ATGATGATGA CAACGACATA ATGGAAACGC TTTGTAAAAA TGATTTGCA 840  
 CTGAAAATAA AAGTGAAGGA GATAACCTAC ATCAACCGAG ATACCAAAAT CATCCTGGAG 900  
 ACCAAGAGCA AGACCATTIA CAAGCTGAAC GGTGTGTCCG AAAGGGACCT GAAGAAATCG 960  
 GTGCTGTGGC TCAAAGACAG CTTGCAGTGC ACCTGTGAGG AGATGAACGA CATCAACGCG 1020  
 CCCTATCTGG TCATGGGACA GAAACAGGGT GGGGAGCTGG TGATCACCTC GGTGAAGCGG 1080  
 TGGCAGAAGG GGCAGAGAGA GTTCAAGCGC ATCTCCCGCA GCATCCGCAA GCTGCAGTGC 1140  
 TAGTCCCGGC ATCCTGATGG CTCCGACAGG CCTGCTCCAG AGCACGGCTG ACCATTCTG 1200  
 CTCCGGGATC TCAGCTCCCG TTCCCCAAGC ACACTCCTAG CTGCTCCAGT CTCAGCCTGG 1260  
 GCAGCTTCCC CCTGCCTTTT GCACGTTGC ATCCCCAGCA TTTCCTGAGT TATAAGGCCA 1320  
 CAGGAGTGGA TAGCTGTTT CACCTAAAGG AAAAGCCCAC CCGAATCTTG TAGAAATATT 1380  
 CAAACTAATA AAATCATGAA TATTTTATG AAGTTAAAAA ATAGCTCACT TTAAAGCTAG 1440  
 TTTTGAATAG GTGCAACTGT GACTTGGTC TGGTTGGTTG TTGTTGTTG TTTTGAGTCA 1500  
 GCTGATTTC ACTTCCCCT GAGGTTGTCA TAACATGCAA ATTGCTTCAA TTTTCTCTGT 1560  
 GGCCCAAAC TGTGGTCAC AAACCCTGTT GAGATAAACG TGGCTGTTAT CTCAACATCT 1620  
 TCATCAGCTC CAGACTGAGA CTCAGTGTCT AAGTCTTACA ACAATTCACT ATTTTATACC 1680  
 TTCAATGGGA ACTTAAACTG TTACATGTAT CACATTCCAG CTACAATACT TCCATTATT 1740  
 AGAAGCACAT TAACCATTTC TATAGCATGA TTTCTTCAAG TAAAAGGCAA AAGATATAAA 1800  
 TTTTATAATT GACTTGAGTA CTTTAAGCCT TGTTTAAAC ATTCTTACT TAACTTTGCA 1860  
 AAATTAAACC CATTGTAGCT TACCTGTAAT ATACATAGTA GTTTACCTTT AAAAGTTGTA 1920  
 AAAATATTGC TTTAACCAAC ACTGTAAATA TTTCAGATAA ACATTATATT CTTGTATATA 1980  
 AACTTTACAT CCTGTTTAC CTAAAAAAAAA AAAAAAAAAG CGGCCGC 2027

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 295 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Leu Gln Gly Pro Gly Ser Leu Leu Leu Phe Leu Ala Ser His  
 1 5 10 15

Cys Cys Leu Gly Ser Ala Arg Gly Leu Phe Leu Phe Gly Gln Pro Asp  
 20 25 30

Phe Ser Tyr Lys Arg Ser Asn Cys Lys Pro Ile Pro Ala Asn Leu Gln  
 35 40 45

Leu Cys His Gly Ile Glu Tyr Gln Asn Met Arg Leu Pro Asn Leu Leu  
 50 55 60

Gly His Glu Thr Met Lys Glu Val Leu Glu Gln Ala Gly Ala Trp Ile  
 65 70 75 80

Pro Leu Val Met Lys Gln Cys His Pro Asp Thr Lys Lys Phe Leu Cys  
 85 90 95

Ser Leu Phe Ala Pro Val Cys Leu Asp Asp Leu Asp Glu Thr Ile Gln  
 100 105 110

Pro Cys His Ser Leu Cys Val Gln Val Lys Asp Arg Cys Ala Pro Val  
 115 120 125

Met Ser Ala Phe Gly Phe Pro Trp Pro Asp Met Leu Glu Cys Asp Arg  
 130 135 140

Phe Pro Gln Asp Asn Asp Leu Cys Ile Pro Leu Ala Ser Ser Asp His  
 145 150 155 160

Leu Leu Pro Ala Thr Glu Glu Ala Pro Lys Val Cys Glu Ala Cys Lys  
 165 170 175

Asn Lys Asn Asp Asp Asp Asn Asp Ile Met Glu Thr Leu Cys Lys Asn  
 180 185 190

Asp Phe Ala Leu Lys Ile Lys Val Lys Glu Ile Thr Tyr Ile Asn Arg  
 195 200 205

Asp Thr Lys Ile Ile Leu Glu Thr Lys Ser Lys Thr Ile Tyr Lys Leu  
 210 215 220

Asn Gly Val Ser Glu Arg Asp Leu Lys Lys Ser Val Leu Trp Leu Lys  
 225 230 235 240

Asp Ser Leu Gln Cys Thr Cys Glu Glu Met Asn Asp Ile Asn Ala Pro  
 245 250 255

Tyr Leu Val Met Gly Gln Lys Gln Gly Gly Glu Leu Val Ile Thr Ser  
 260 265 270

Val Lys Arg Trp Gln Lys Gly Gln Arg Glu Phe Lys Arg Ile Ser Arg  
 275 280 285

Ser Ile Arg Lys Leu Gln Cys  
 290 295

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 275 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Ser Ala Arg Gly Leu Phe Leu Phe Gly Gln Pro Asp Phe Ser Tyr Lys  
1 5 10 15

Arg Ser Asn Cys Lys Pro Ile Pro Ala Asn Leu Gln Leu Cys His Gly  
20 25 30

Ile Glu Tyr Gln Asn Met Arg Leu Pro Asn Leu Leu Gly His Glu Thr  
35 40 45

Met Lys Glu Val Leu Glu Gln Ala Gly Ala Trp Ile Pro Leu Val Met  
50 55 60

Lys Gln Cys His Pro Asp Thr Lys Lys Phe Leu Cys Ser Leu Phe Ala  
65 70 75 80

Pro Val Cys Leu Asp Asp Leu Asp Glu Thr Ile Gln Pro Cys His Ser  
85 90 95

Leu Cys Val Gln Val Lys Asp Arg Cys Ala Pro Val Met Ser Ala Phe  
100 105 110

Gly Phe Pro Trp Pro Asp Met Leu Glu Cys Asp Arg Phe Pro Gln Asp  
115 120 125

Asn Asp Leu Cys Ile Pro Leu Ala Ser Ser Asp His Leu Leu Pro Ala  
130 135 140

Thr Glu Glu Ala Pro Lys Val Cys Glu Ala Cys Lys Asn Lys Asn Asp  
145 150 155 160

Asp Asp Asn Asp Ile Met Glu Thr Leu Cys Lys Asn Asp Phe Ala Leu  
165 170 175

Lys Ile Lys Val Lys Glu Ile Thr Tyr Ile Asn Arg Asp Thr Lys Ile  
180 185 190

Ile Leu Glu Thr Lys Ser Lys Thr Ile Tyr Lys Leu Asn Gly Val Ser  
195 200 205

Glu Arg Asp Leu Lys Lys Ser Val Leu Trp Leu Lys Asp Ser Leu Gln  
210 215 220

Cys Thr Cys Glu Glu Met Asn Asp Ile Asn Ala Pro Tyr Leu Val Met  
225 230 235 240

Gly Gln Lys Gln Gly Gly Glu Leu Val Ile Thr Ser Val Lys Arg Trp  
245 250 255

Gln Lys Gly Gln Arg Glu Phe Lys Arg Ile Ser Arg Ser Ile Arg Lys  
260 265 270

Leu Gln Cys  
275

*MW 02*